DOCUMENT RESUME

ED 068 579 TM 002 094

AUTHOR Pettibone, Timothy J.; Diamond, James J.
TITLE An Incorrect Index of Skewness.
PUB DATE [72]
NOTE 8p.

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS Correlation; *Mathematical Models; Measurement Techniques; Methodology; Nonparametric Statistics; *Reliability; Research Methodology; *Statistical Analysis; Test Construction; *Test Validity

ABSTRACT

Checking of parametric assumptions is an often ignored step in the inferential process. A misconception regarding symmetry (one aspect of "robustness") is prevalent: that in nonsymmetric distributions the mean and median are always non-coincidental. It is to this fallacious point that the discussion is directed. A generally accepted index of skewness, as identified in mathematical statistics texts, is the third moment of a random variable around its mean, but this index, sometimes denoted as Beta , is really the average cubed linear z score for a distribution. Thus if the distribution is symmetric, Beta will be zero. By considering two nonsymmetric distributions, further discussion of the point is given. In conclusion, when there is concern about symmetry (for either pedagogical or technical reasons), the data should be plotted and the distribution examined. This inspection technique along with Beta computation will result in a more valid conclusion regarding symmetry. (LH)

ED 068579

TM 002 094

1

AN INCORRECT INDEX OF SKEWNESS

Timothy J. Pettibone
New Mexico State University

James J. Diamond
University of Pennsylvania

Checking of parametric assumptions is an often ignored step in the inferential process. Typically, experimental data are gathered and a statistical test is applied. This test will probably have been drawn from those known as "robust." Boneau (1960) stated: that provided "the statistical test is relatively insensitive to violations of the assumptions other than the null hypothesis, and, hence, if probability statements refer primarily to the null hypothesis, it is said to be robust. The t and F tests apparently possess this quality to a high degree (p. 63)." On the other hand, Bradley (1968) talks of the "myth of robustness." To him, " ... a kernel of truth has been magnified into a mountain of error (p. 24)." A middle position regarding robustness has been taken by Lindquist (1953). For example, according to Lindquist's point of view, although the t test is robust, the robustness depends to a great extent on distribution form and variance.

Whatever position the reader holds, a misconception regarding symmetry (one aspect of robustness) has been around for a long time and continues to exist in the current literature. This is the well-known (but entirely erroneous) "fact" that in nonsymmetric distributions the mean and median are always non-coincidental ($\bar{X} - P_{50} \neq 0$). It is to this fallacious point that this discussion is directed.

A generally accepted index of skewness, as identified in mathematical statistics texts, is the third moment of a random variable around its mean. For example, see Weatherburn (1962) and Wilks (1962). This index, sometimes denoted as β_1, is really the average cubed linear z score for a distribution. That is,

$$\beta_1 = \frac{\sum_{i=1}^{n} z_i^3}{n} .$$

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

Supposedly, if the distribution is symmetric β_1 will be zero. In addition, β_1 supposedly will be positive or negative for positively or negatively skewed distributions respectively.

The simpler index of skewness ($\bar{X} - P_{50}$) has a rather specific property which needs to be considered in interpreting this seemingly parsimonious condition for symmetry. That is, $\bar{X} - P_{50} = 0$ can be obtained for a nonsymmetric distribution, a point not mentioned in many introductory texts (Ary, Jacobs and Razavich, 1972); Englehart, 1972; Ferguson, 1971, Freund, 1967; Garrett, 1966; Glass and Stanley, 1970; Hill and Kerber, 1967; Senter, 1969; Walker and Lev, 1958; Wert, Neidt and Ahmann, 1954). As Ary, et al (1972) noted: "If a distribution of measures is symmetrical, the values of the mean and the median coincide ... . If a distribution is not symmetrical ... the values of the measures of central tendency differ (p. 104)." The first part is absolutely correct. In a symmetric distribution $\bar{X} = P_{50}$. However, this is a necessary but not a sufficient condition for symmetry. When these same authors (p. 105) write: "The skew of a distribution can be identified by comparing the mean and the median without necessarily constructing a histogram or polygon," it shows, in explicit form, the common error in logic.

To be completely fair, the point has been mentioned by Hays (1963) when he says: "A word of warning: if a distribution is symmetric, then Mean = Median, but the fact the Mean = Median does not necessarily imply that the distribution is symmetric (p. 174)."

Some important lessons, apparently in need of review, can be obtained from a look at this phenomenon. Let us consider a distribution of some undefined shape based upon N scores. For simplicity, consider N to be even. By definition, there will be N/2 scores both below and above P_{50}. For this point also to be $\bar{X}$ it is necessary that certain sums be equal. Consider the absolute value of the deviation

of each score X_i from the mean $\bar{X}$. In order for $\bar{X}$ and P_{50} to be the same point it is necessary that

$$\sum_{i=1}^{N/2} \left|(X_i - \bar{X})\right| = \sum_{i=N/2+1}^{N} \left|(X_i - \bar{X})\right|$$

Notice that we do not require the separate deviations to be equal, but only their sum. That is, it is not necessary for the absolute value of $(X_1 - \bar{X})$ to equal $(X_N - \bar{X})$, and $(X_2 - \bar{X})$ to equal $(X_{N-1} - \bar{X})$ and so on, as they would in a symmetric distribution. Therefore it is possible for $\bar{X}$ and P_{50} to be equal in a non-symmetric distribution. One of the many possible illustrations is shown in Figure 1. For this distribution $\bar{X}$ and P_{50} equal 6 but the distribution is certainly not symmetric. Note, however, that β_1 does not equal zero but signifies a negatively skewed distribution. (Assymetric distributions can be constructed where β_1 may itself be zero but this seems to be a rare phenomenon.)

- - - - - - - - - - - - - - -
Insert Figure 1 about here
- - - - - - - - - - - - - - -

A colleague criticizing a preliminary draft of this paper pointed out that our unusual distribution wasn't particularly convincing and asked for a more conventional nonsymmetric distribution. Figure 2 depicts a more typical distribution showing a lack of symmetry. As before $\bar{X} - P_{50} = 0$ and $\beta_1 \neq 0$.

- - - - - - - - - - - - - - -
Insert Figure 2 about here
- - - - - - - - - - - - - - -

One may ask, "why all this concern about lack of symmetry?" An obvious answer, of course, is that the $\bar{X} - P_{50}$ fallacy is widespreadly held and taught. Another answer is in regard to violation of parametric assumptions. Lindquist (1953) stated that, " ... unless heterogeneity of either form or variance is so extreme as to be readily apparent upon inspection of the data, the effect upon the F distribution will probably be negligible (p. 86)."

In conclusion then, if one is concerned about symmetry (for either pedagogical or technical reasons), one should at least plot the data and look at the distribution. This inspection technique along with β_1 computation, will result in a more valid conclusion regarding symmetry.

References

Ary, D., Jacobs, L. C., & Razavich, A. Introduction to research in education. New York: Holt, 1972.

Boneau, C. A. The effects of violations of assumptions underlying the t test. Psychological Bulletin, 1960, 57, (1), 49-64.

Bradley, J. V. Distribution-free statistical tests. Englewood Cliffs: Prentice-Hall, 1968.

Englehart, M. D. Methods of educational research, New York: McGraw-Hill, 1972.

Ferguson, G. A. Statistical analysis in psychology and education. New York: McGraw-Hill, 1971.

Freund, J. E. Modern elementary statistics (3rd. ed.). Englewood Cliffs: Prentice-Hall, 1967.

Garrett, H. E. Statistics in psychology and education (6th. ed.). New York: David McKay, 1966.

Glass, G. V , & Stanley, J. Statistical methods in education and psychology. Englewood Cliffs: Prentice-Hall, 1970.

Hays, W. L. Statistics for psychologists. New York: Holt, 1963.

Hill, J. E., & Kerber, A. Models, methods and analytical procedures in education research. Detroit: Wayne State University Press, 1967.

Lindquist, E. F. Design and analysis of experiments in psychology and education. Boston: Houghton-Mifflin, 1953.

Senter, R. J. Analysis of data. Glenview, Illinois: Scott, Foresman and Co., 1969.

Walker, H. M., & Lev, J. Elementary statistical methods. New York: Holt, 1958.

Weatherburn, C. E. A first course in mathematical statistics. Cambridge: Cambridge University Press, 1962.

Wert, J. E., Neidt, C. O., & Ahmann, J. S. Statistical methods in education and psychological research. New York: Appleton-Century-Crofts, 1954.

Wilks, S. S. Mathematical statistics. New York: Wiley, 1962.

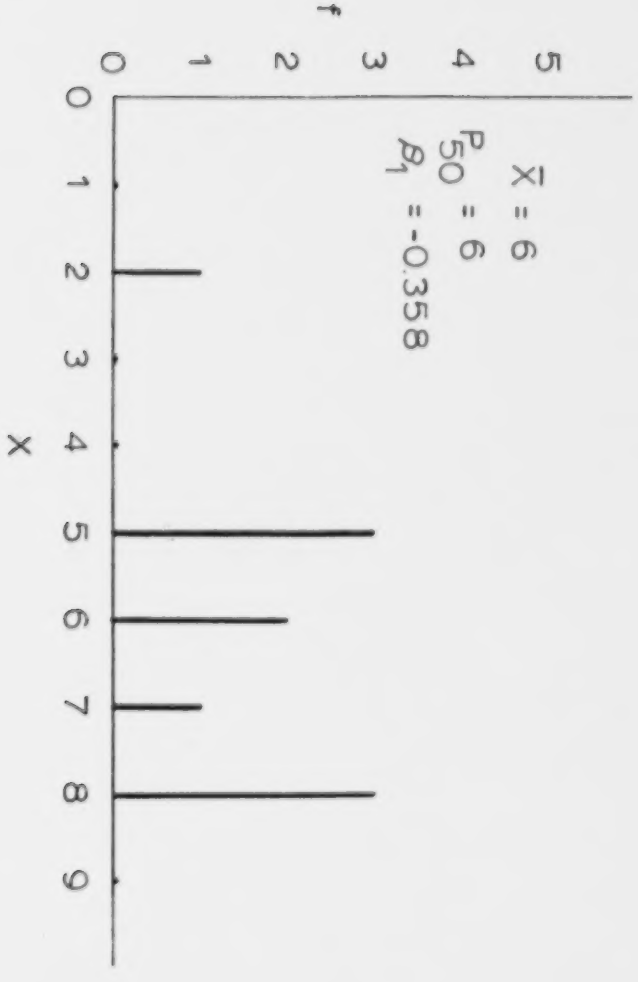
Figure Captions

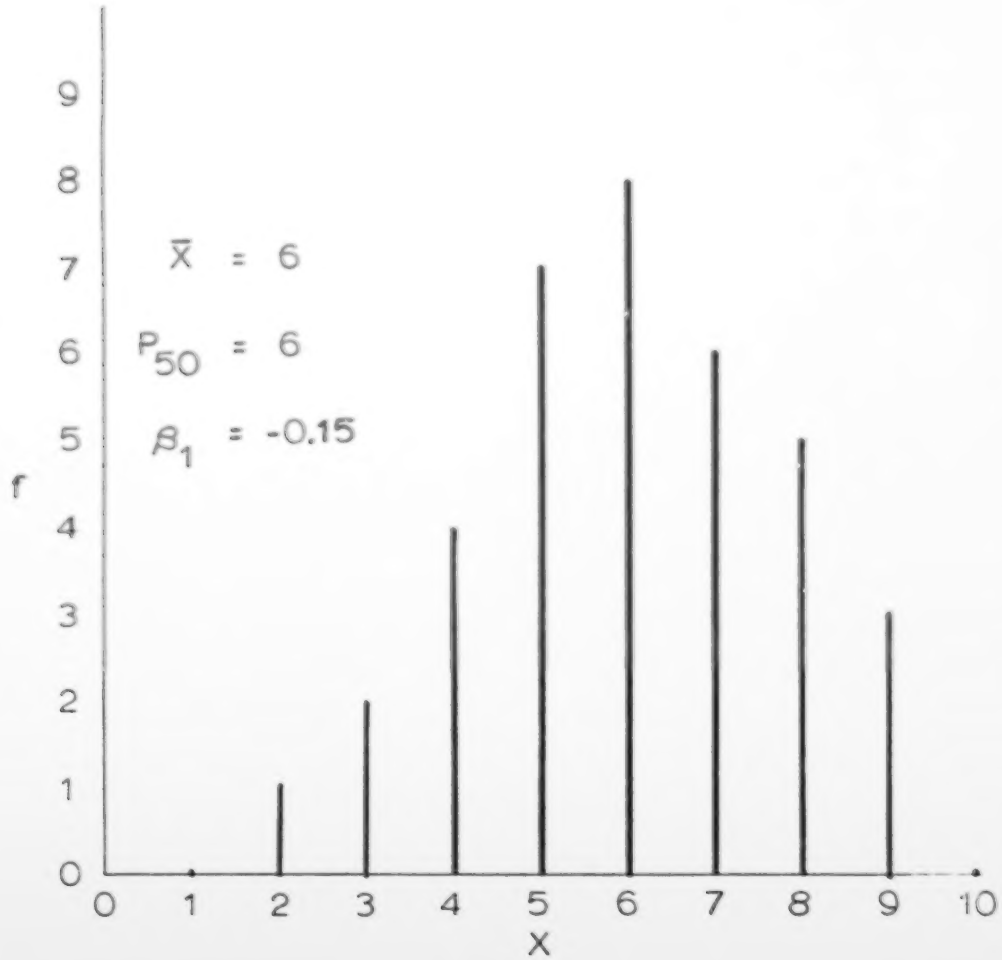
Page 7

Figure 1, A nonsymmetric distribution where $\bar{X} - P_{50} = 0$.

Page 8

Figure 2, A more typical nonsymmetric distribution


f
0 1 2 3 4 5
0 1 2 3 4 5 6 7 8 9
X
$\bar{X} = 6$
$P_{50} = 6$
$\beta_1 = -0.358$


$\bar{X}$ = 6
P_{50} = 6
β_1 = -0.15
f
X
0 1 2 3 4 5 6 7 8 9 10
0 1 2 3 4 5 6 7 8 9